

Artificial General Intelligence

AI-PHI

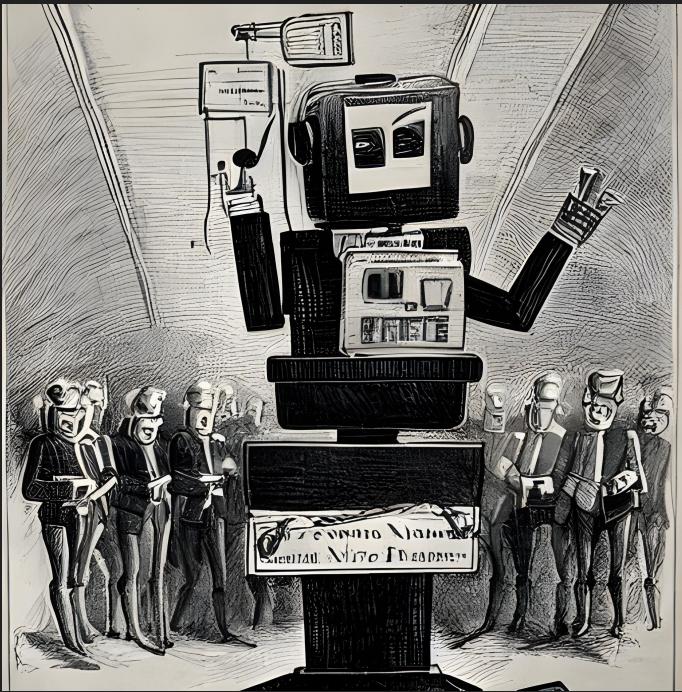
Join our WhatsApp Group



Ai-Phi



AI News - Hot of the Transformer!



ToC

- A DENSE overview of AI
- AI Trials and Tribulations
- AGI as a ‘Field’
- ChatGPT

Rough Areas of AI - According to Russell and Norvig 2022

KR & Reasoning

Search

Uncertainty

I Artificial Intelligence

- 1 Introduction ... 1
- 2 Intelligent Agents ... 36

II Problem-solving

- 3 Solving Problems by Searching ... 63
- 4 Search in Complex Environments ... 110
- 5 Adversarial Search and Games ... 146
- 6 Constraint Satisfaction Problems ... 180

III Knowledge, reasoning, and planning

- 7 Logical Agents ... 208
- 8 First-Order Logic ... 251
- 9 Inference in First-Order Logic ... 280
- 10 Knowledge Representation ... 314
- 11 Automated Planning ... 344

IV Uncertain knowledge and reasoning

- 12 Quantifying Uncertainty ... 385
- 13 Probabilistic Reasoning ... 412
- 14 Probabilistic Reasoning over Time ... 461
- 15 Probabilistic Programming ... 500
- 16 Making Simple Decisions ... 528
- 17 Making Complex Decisions ... 562
- 18 Multiagent Decision Making ... 599

Machine Learning

Perception | Data

V Machine Learning

- 19 Learning from Examples ... 651
- 20 Learning Probabilistic Models ... 721
- 21 Deep Learning ... 750
- 22 Reinforcement Learning ... 789

VI Communicating, perceiving, and acting

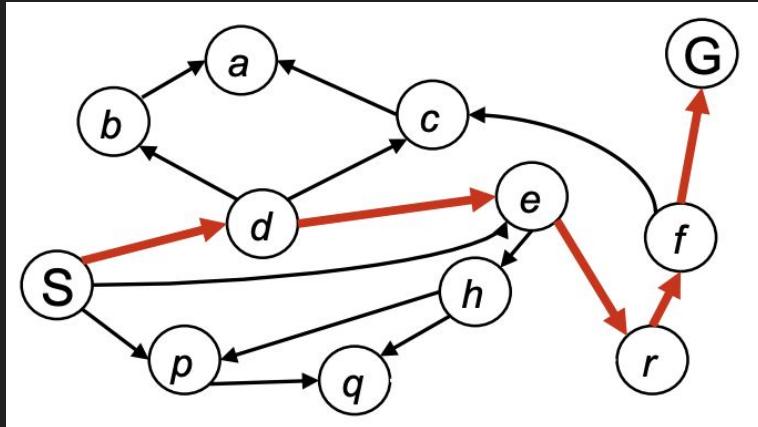
- 23 Natural Language Processing ... 823
- 24 Deep Learning for Natural Language Processing ... 856
- 25 Computer Vision ... 881
- 26 Robotics ... 925

VII Conclusions

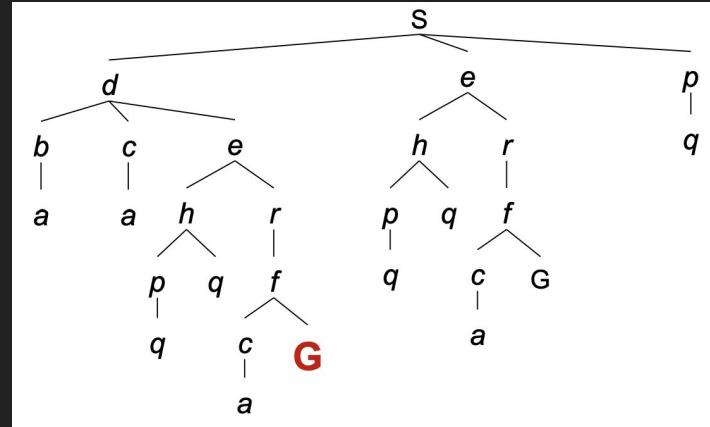
- 27 Philosophy, Ethics, and Safety of AI ... 981
- 28 The Future of AI ... 1012
- Appendix A: Mathematical Background ... 1023
- Appendix B: Notes on Languages and

Search

State Space



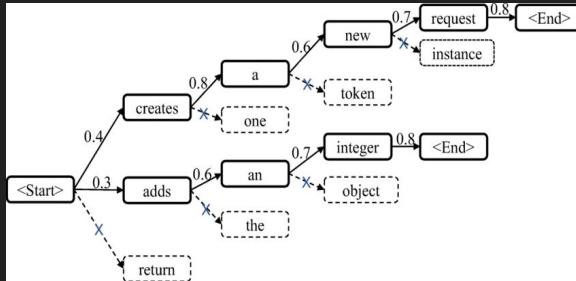
Search Tree



Moving through a state space to maximise or minimise a goal

Search - Examples

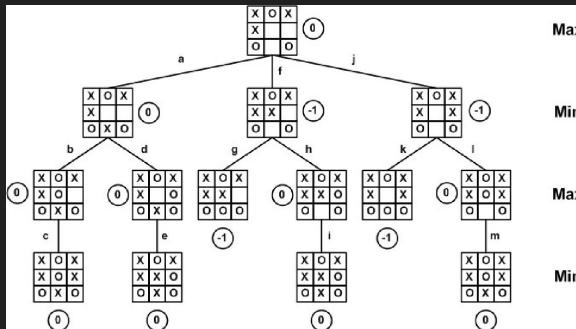
Beam Search Decoding



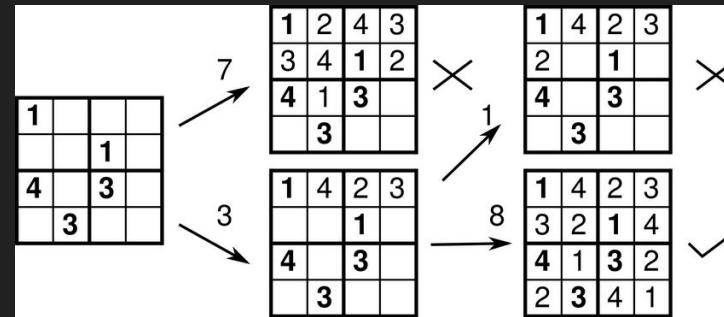
Shortest Path



Adversarial Search

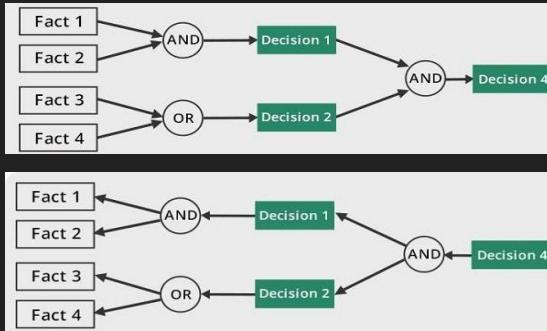


Solving Sudoku



KR & Reasoning - Certainty

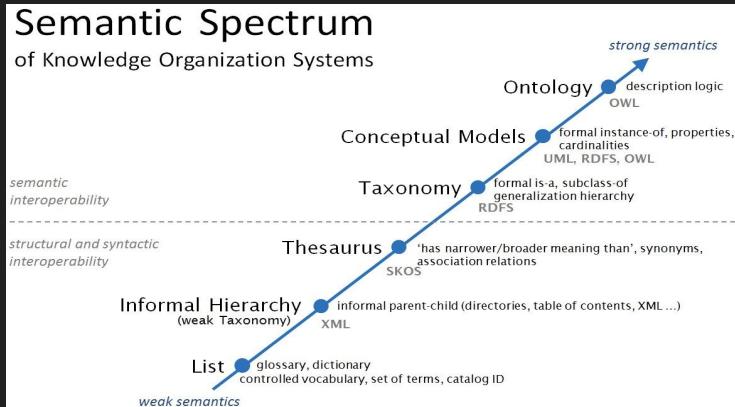
Forward & Backward Chaining



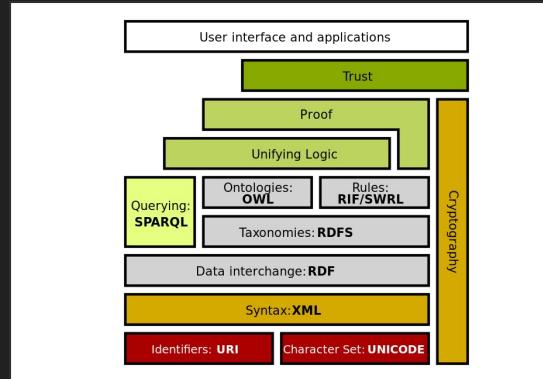
Structure of Knowledge

Semantic Spectrum

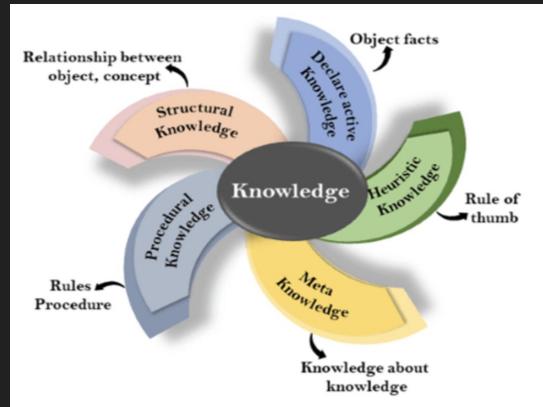
of Knowledge Organization Systems



Semantic Web

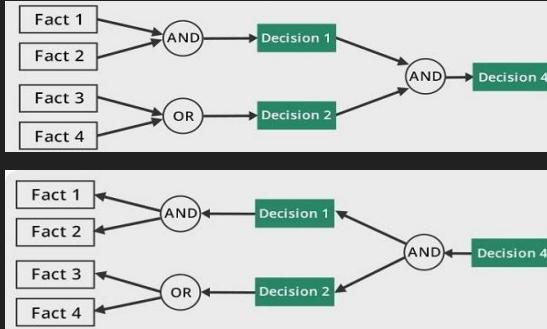


Types of Knowledge



KR & Reasoning - Certainty

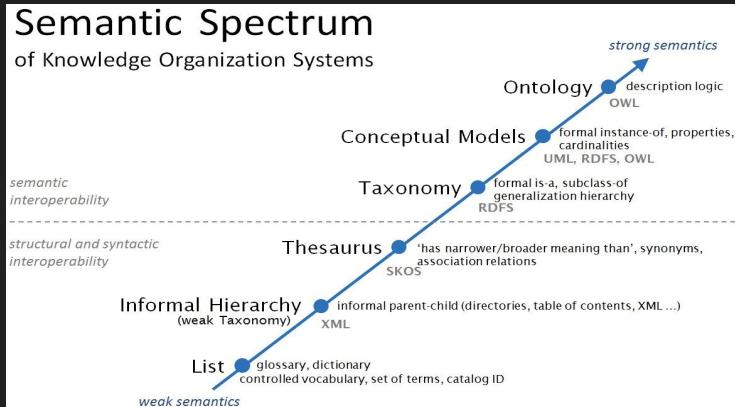
Forward & Backward Chaining



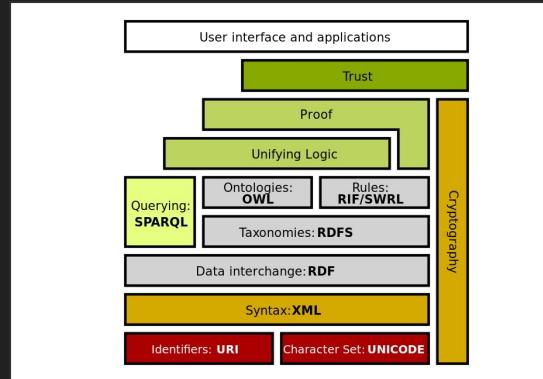
Structure of Knowledge

Semantic Spectrum

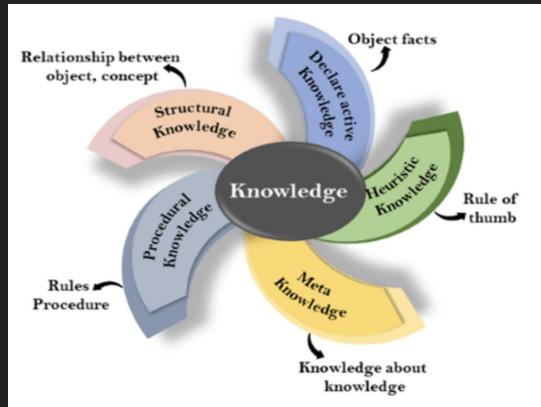
of Knowledge Organization Systems



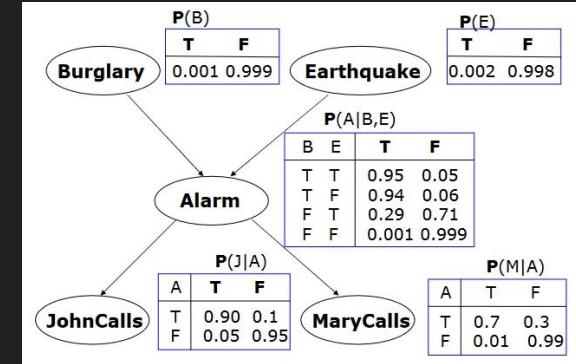
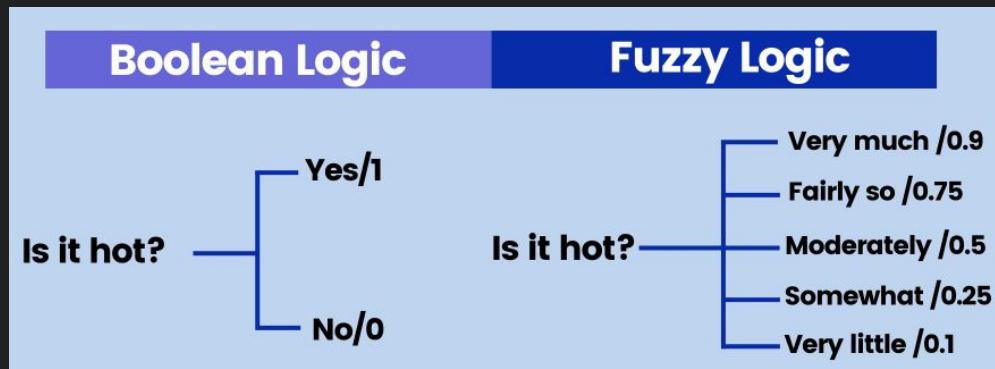
Semantic Web



Types of Knowledge

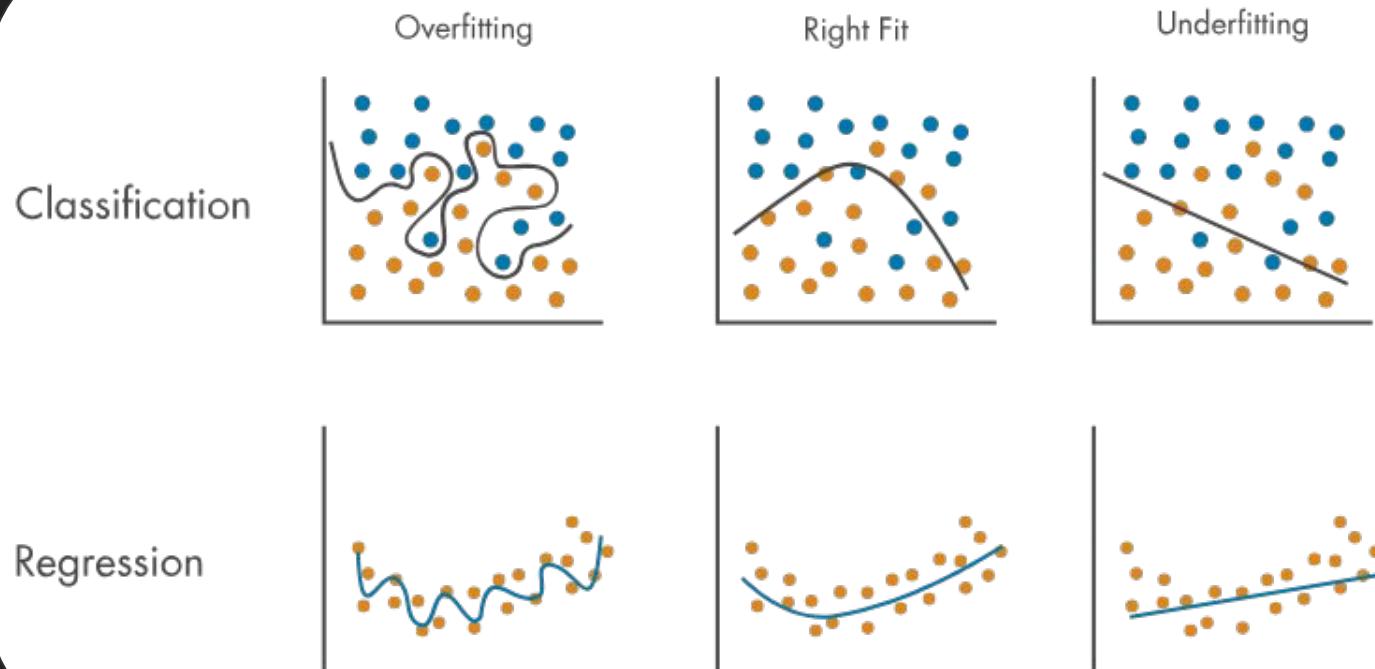


KR & Reasoning - Uncertainty



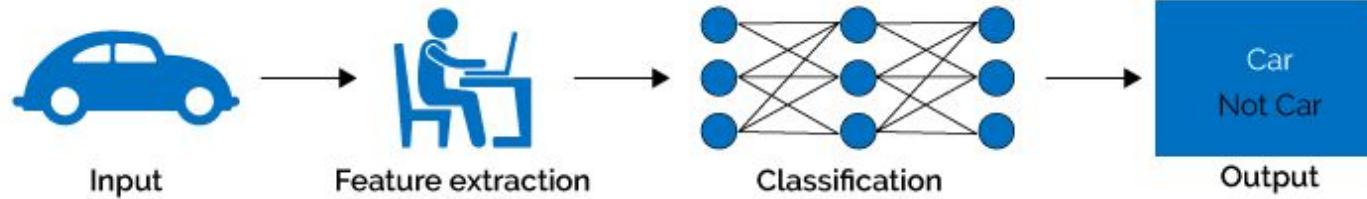
Natural Language Processing

Machine Learning

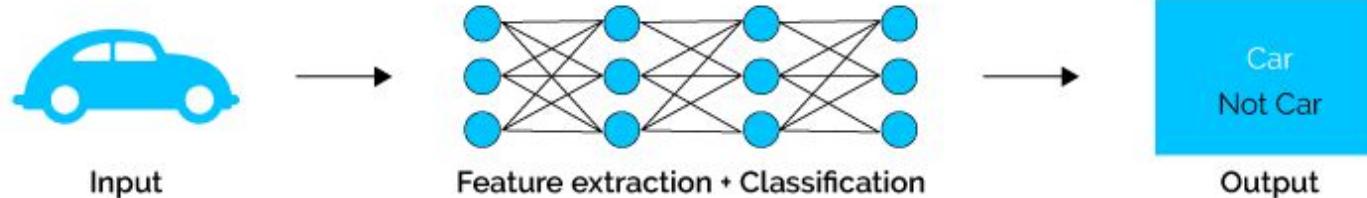


Machine Learning & Deep Learning

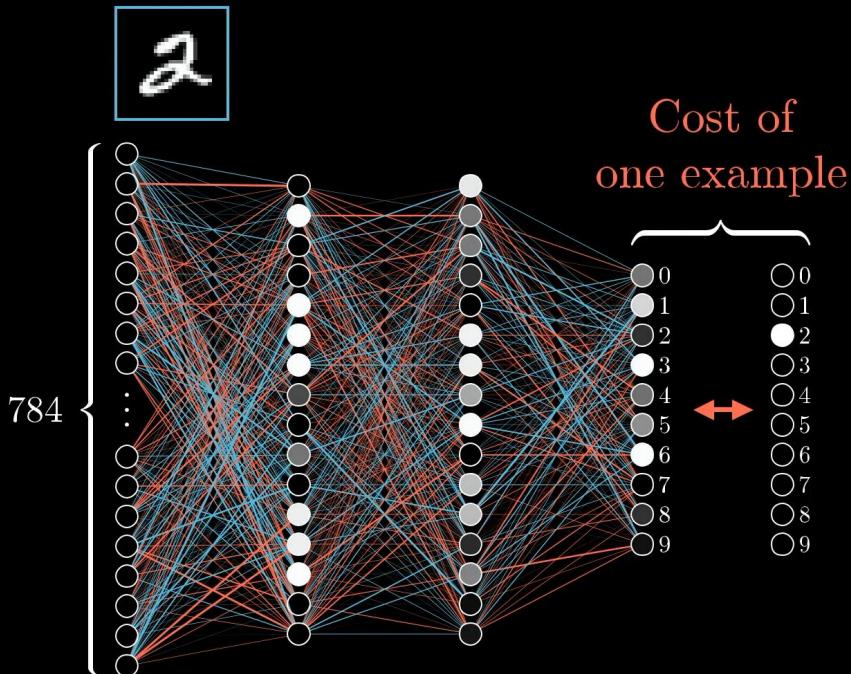
Machine Learning



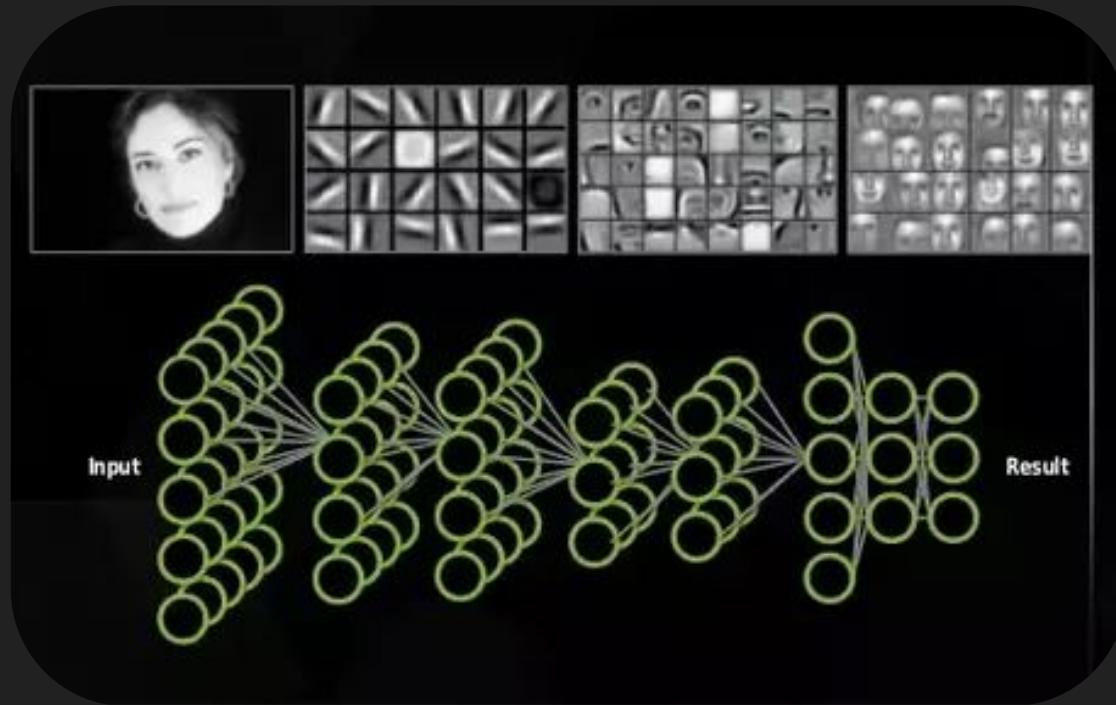
Deep Learning



Neural Networks - Backpropagation

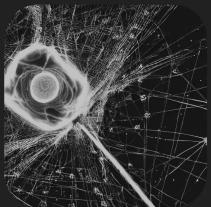


Neural Networks - Feature Learning



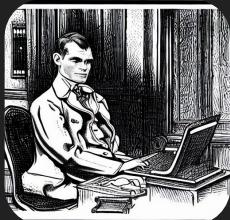
AI Milestones

Artificial Neurons
McCulloch & Pitts



1943

Turing Test



2020

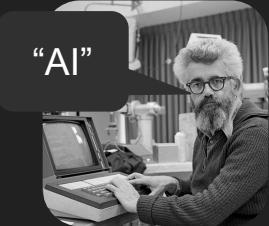


First Impressive
Language Model
Deep Learning



IBM Watson
Wins Jeopardy
NLP + KR & Reasoning + Search

Dartmouth Conference
John McCarthy



1954

Early NLP

```
EEEEEELL    IIII    ZZZZZZ  AAAAAA
EEEEEELL    II     ZZ     AA     AA
EEEEEELL    II     ZZZ  AAAAAAA
EEEEEELL    II     ZZ     AA     AA
EEEEEELLLL  IIII    ZZZZZZ  AA     AA
```

A is a mock Bopprean psychopathologist.
original program was described by Joseph Weizenbaum in 1966
implementation by Norbert Landsteiner 2005.

Is something troubling you ?
Men are all alike.
What's your occupation, do you suppose ?
They're always bugging us about something or other.
Can you think of a specific example ?
Well, my boyfriend made me come here.
He says I'm depressed much of the time.
I am sorry to hear that you are depressed.
It's true. I am unhappy.
an you explain what made you unhappy ?

First
Chatbot

1966



1997

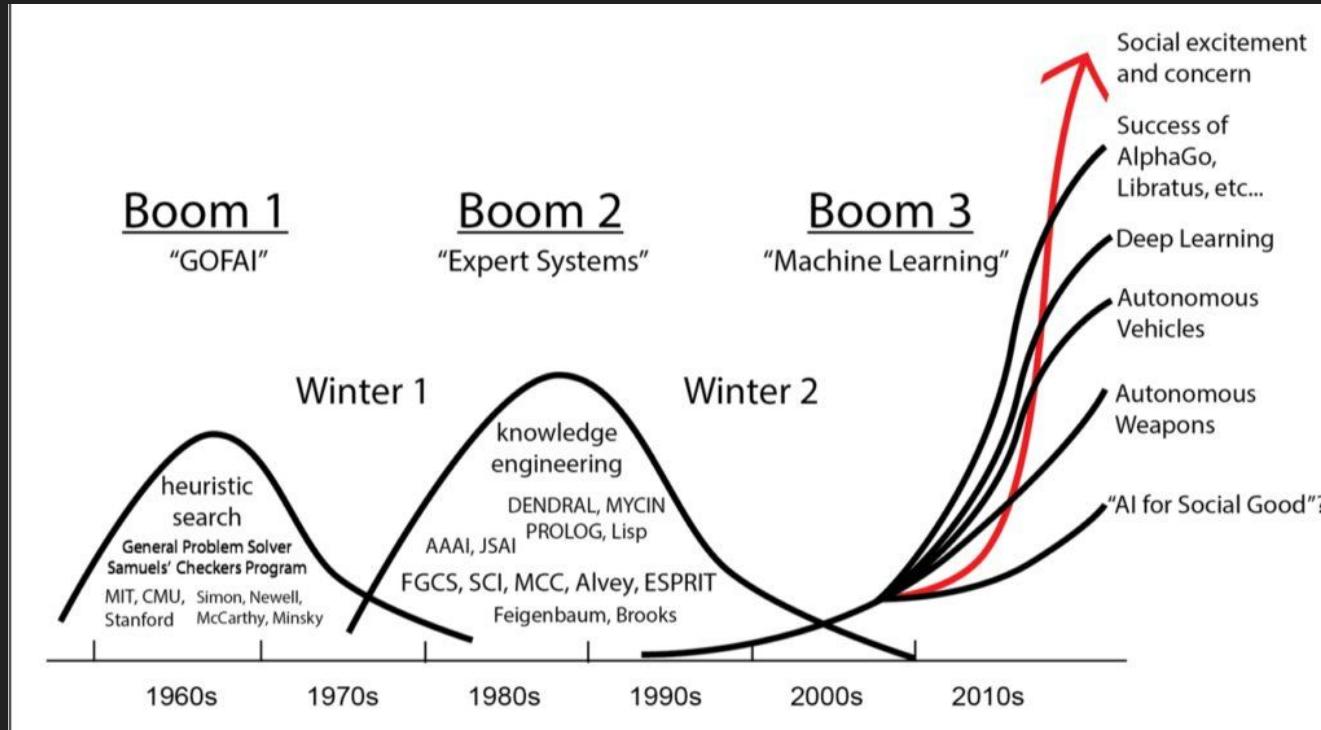
KR & Reasoning



Deep Blue
Beats Kasparov
Adversarial Search

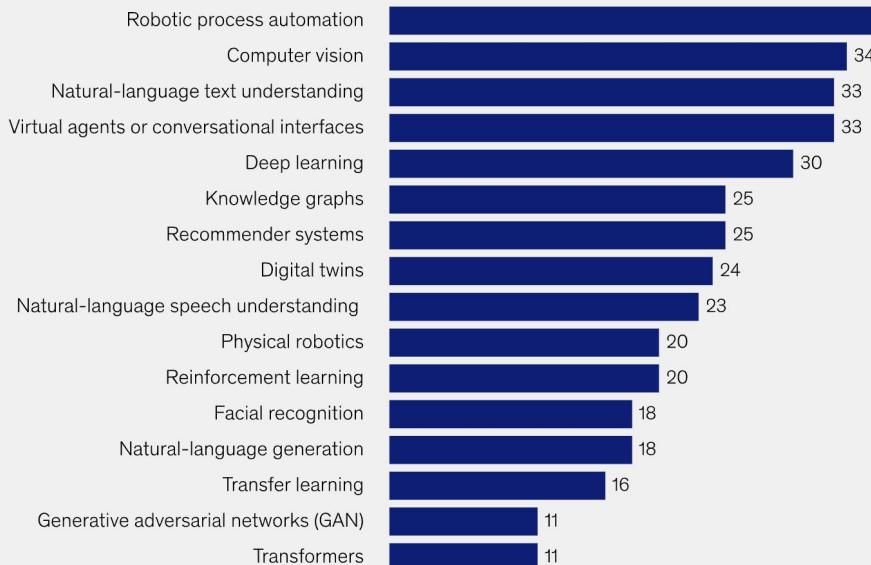
Dendral
Expert System

AI Winters & Hype



In Industry

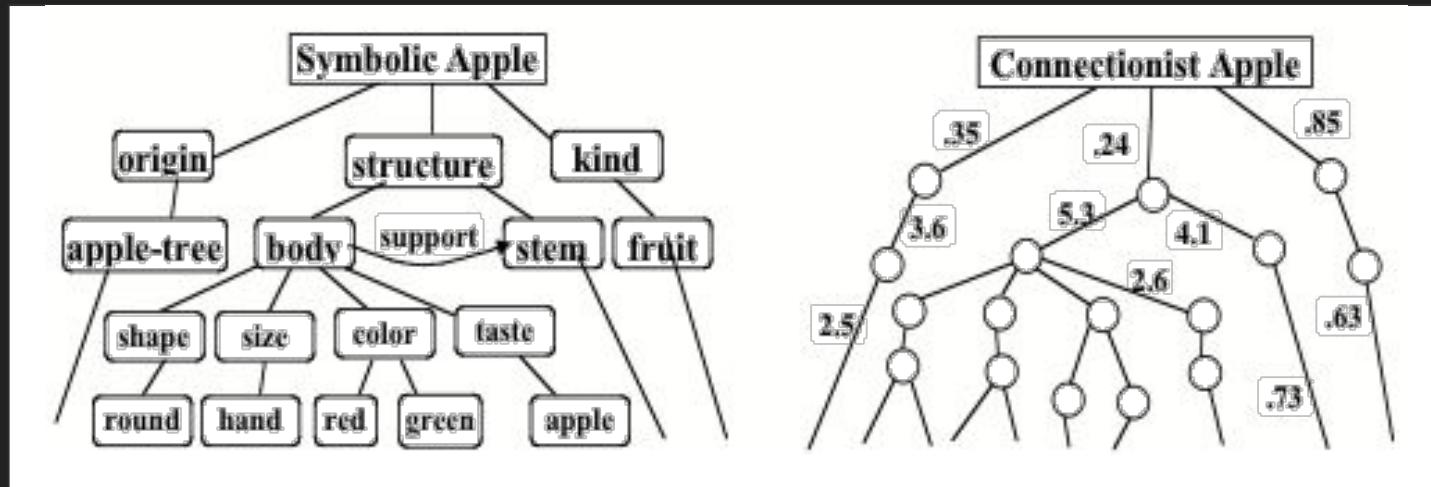
Percentage of respondents who say given AI capability is embedded in products or business processes in at least one function or business unit²



¹The number of capabilities included in the survey has grown over time, from 9 in 2018 to 15 in the 2022 survey.

²Question was asked only of respondents who said their organizations have adopted AI in at least one function.

Tensions in AI - Symbolist vs Connectionist



Tensions in AI - Domain Expertise vs

- *Symbolist vs connectionist vs hybrid*
- *Understanding problems vs solving them*
 - *Explainability vs pragmatism*

The AGI Hypothesis

“the creation and study of synthetic intelligences with sufficiently broad (e.g. human-level) scope and strong generalization capability, is at bottom qualitatively different from the creation and study of synthetic intelligences with significantly narrower scope and weaker generalization capability.”

Goertzel, Ben. "Artificial General Intelligence: Concept, State of the Art, and Future Prospects" Journal of Artificial General Intelligence, vol.5, no.1, 2014, pp.1-48. <https://doi.org/10.2478/jagi-2014-0001>

The AGI Hypothesis

“the creation and study of synthetic intelligences with sufficiently broad (e.g. human-level) scope and strong generalization capability, is at bottom qualitatively different from the creation and study of synthetic intelligences with significantly narrower scope and weaker generalization capability.”

AGI (Science Fiction)

Broad AI (IBM)

Narrow AI

Goertzel, Ben. "Artificial General Intelligence: Concept, State of the Art, and Future Prospects" Journal of Artificial General Intelligence, vol.5, no.1, 2014, pp.1-48. <https://doi.org/10.2478/jagi-2014-0001>

AGI - Broad Definition

- AI that achieves a variety of goals, and carries out a variety of tasks, in a variety of different contexts and environments
- Solve problems that are not in the minds of their creators
- Generalising knowledge and reasoning across contexts
- NOT necessarily just human-level intelligence
- NOT necessarily existing in typical human domains
- NOT necessarily completely without bias
- NOT necessarily arbitrarily good at every possible thing

Goertzel, Ben. "Artificial General Intelligence: Concept, State of the Art, and Future Prospects" Journal of Artificial General Intelligence, vol.5, no.1, 2014, pp.1-48. <https://doi.org/10.2478/jagi-2014-0001>

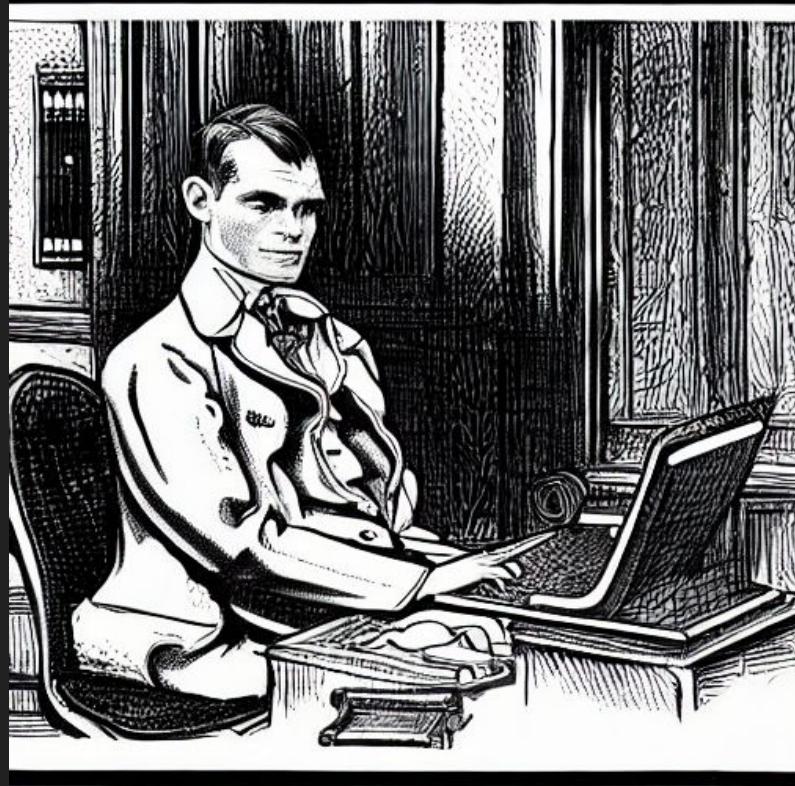
AGI - Humans and AGIs

- We would probably agree that we are generally intelligent
- Any AGI system should overlap (**in some way**) with human intelligence
 - Perhaps having a language for thought or communication?
 - Perhaps explaining (to itself or to us) why it does what it does?
 - Perhaps having motivations and trying to fulfil them?

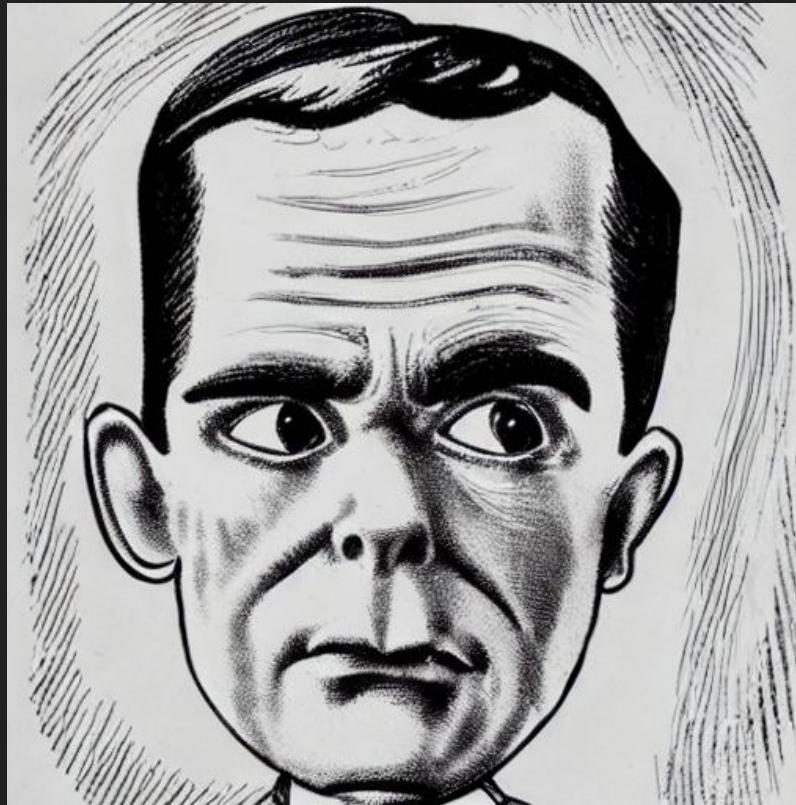
AGI - Non-Human Intelligence

- Hypothetical aliens
 - Might not have all of the systems we have
 - E.g. ‘emotions’ might not be the same
 - Might have different perceptual needs
- Any AGI system should overlap (**in some way**) with hypothetical aliens
 - Perhaps pattern recognition?
 - Perhaps goal oriented?
 - Perhaps consciousness?

The Turing Test



What is this glowing square and this flat typewriter?



The Turing Test

Them: Hi, I'm Vince. Glad to make your acquaintance.
What's your name?

You: ...

Them: I just got out of Beddington. I'm not really feeling like talking. What is it you want?

You: ...

Unknown Interlocutor A

Unknown Interlocutor B

The Turing Test



Monsieur I. A

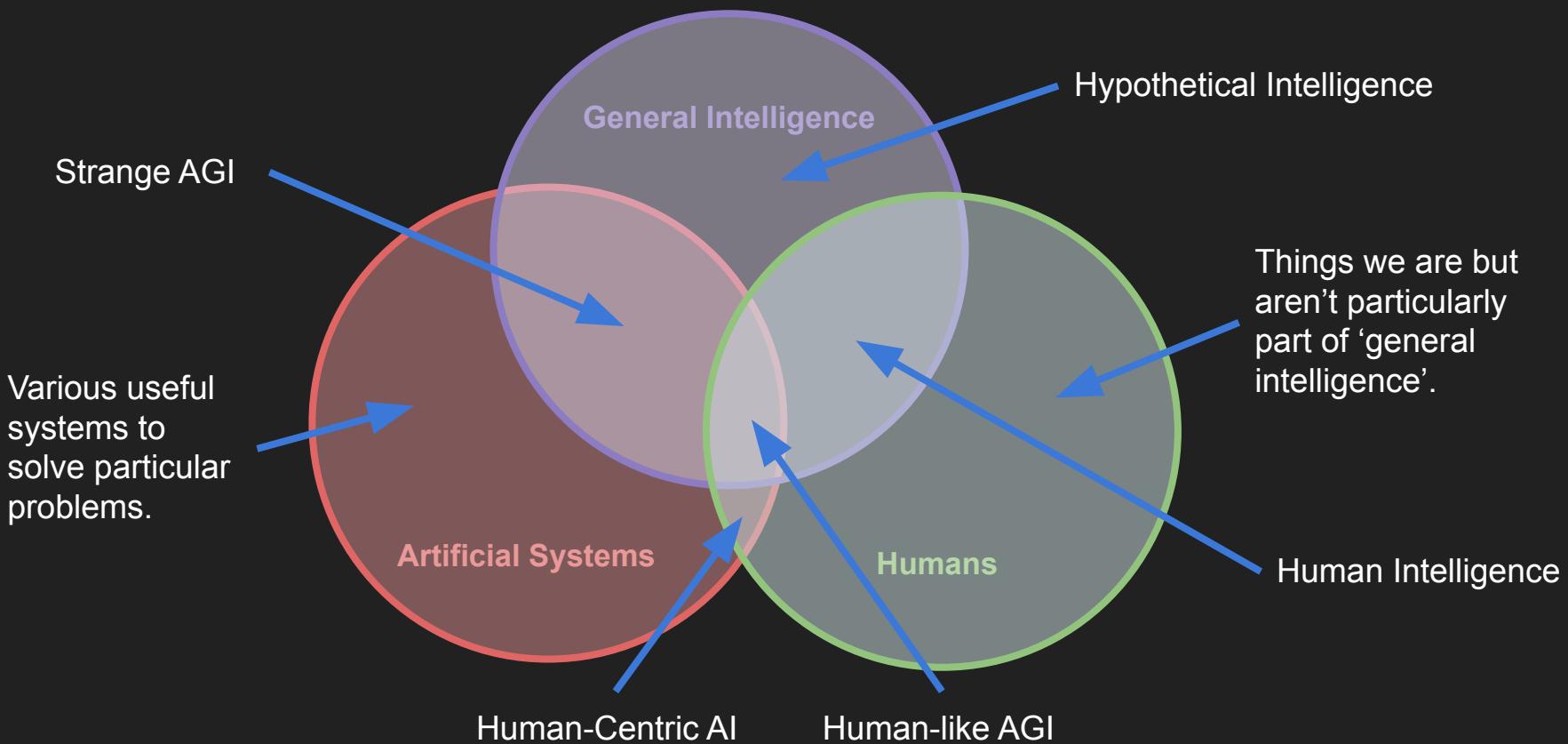


Mister Hugh Man

The Turing Test

- Emulating human text doesn't encompass all of AGI
- If motivations, desires and intentions are important for an AGI, a chatbot can pass the Turing test without 'caring' about passing the test or being able to 'care' at all.

Overlaps with ‘General Intelligence’



A broad set of approaches characterizing GI

- Comparison to Human Competencies
 - Perception, actuation, memory, learning, reasoning, planning, attention, motivation, emotion, modelling oneself and others, social interactions, communication, quantitative, building/creation ...
- A Cognitive-Architecture Perspective
 - Identify fundamental cognitive processes in being intelligent
 - The SOAR architecture (<https://soar.eecs.umich.edu/>)
- A Formal Approach
 - What is intelligence as an abstract concept?
 - E.g. reward-achieving capability and modelling probability distributions over states
- The Adaptationist Approach
 - AGI should be able to adapt to its environment

A broad set of approaches characterizing GI

- The Embodiment Approach
 - REQUIRES embodiment in a physical environment to be generally intelligent
- Pragmatist Approach to GI
 - GI is simply performing well in many contexts and domains
 - Doesn't matter what the architecture or approach is

“... that achieving real Human Level artificial intelligence would necessarily imply that most of the tasks that humans perform for pay could be automated.”

- “Human Level Artificial Intelligence? Be Serious!” (Nilsson, 2005)

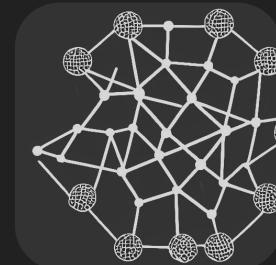
The Field of AGI

- **Symbolic (GOFAI)**
 - Discrete symbols for representation and reasoning
 - Corresponding roughly to units of (human) thought
- **Emergentist**
 - Sub-symbolic representations learned bottom up from data
- **Hybrid**
 - Combining symbolic and emergent systems
- **Universalist**
 - Algorithms, that if computable (but aren't), would probably result in powerful AGI
 - Scale them down until feasible



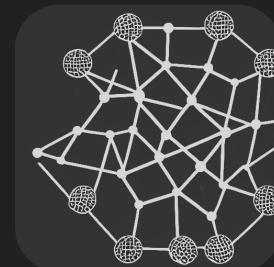
Symbolic Approaches

- Symbolic representation
 - From the mind to the machine
 - Ontologies describing high-level concepts and relationships
 - Knowledge bases representing facts about a domain
 - Structured in various ways such as graphs and logical formulas
- Diverse reasoning approaches
 - Logical reasoning, automated theorem proving, symbolic planning, sat solving, expert systems, multi-agent systems, search algorithms etc.
- The main paradigm of AI from the 1950s-1990s.
 - Sometimes referred to as good old-fashioned AI (GOFAI)



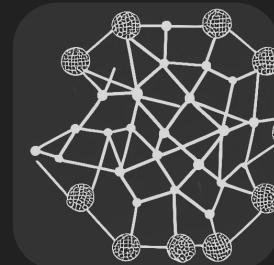
Symbolic - For & Against

- We emit symbols and we are intelligent
 - Perhaps they should also be part of an AGI system?
- Abstract reasoning can occur independently from sensori-motor signals
 - Don't need to **imagine** about all extensions of a category to reason about it
- Symbolic rules can potentially allow for compact and complex reasoning
 - Various mature and diverse approaches to reasoning
- Can infer new knowledge
 - Reasoning can be used to deduce new facts that can be added to our knowledge



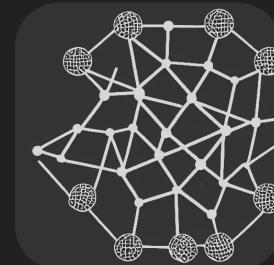
Symbolic - For & Against

- Lies on a spectrum from computationally intractable to trivially computable
- Symbols are typically detached from their meaning
 - Humans have to organise them into systems that perform tasks
 - Symbolic representations are meaningless without humans to interpret them
- The knowledge acquisition bottleneck
 - Even simple domains are complex and can be modelled at arbitrary levels of detail
 - Alignment between independently constructed knowledge can be difficult



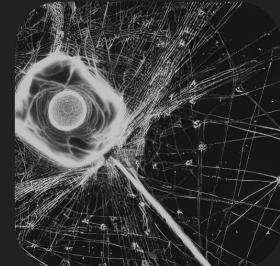
Symbolic - For & Against

- Tend to be very brittle
 - Break when used outside of the precise expectations of their creator
- Monotonic systems
 - New knowledge can't undo old knowledge, once it's there, it's there
- Limited by our own understanding
 - Projecting our knowledge onto a system might limit what it can represent and how it can reason
 - Can't construct what we don't/ can't understand



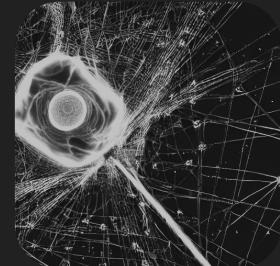
Emergentist (focus on neural networks)

- Sub-symbolic representations
 - Typically vectors and matrices
- ‘Learn’ representations bottom-up from features of the data
- Reasoning is typically inductive:
 - Pattern recognition, prediction, forecasting etc.
- Around since the 1950s.
- Neural networks largely only worked for smallish toy problems until the 2010s with the advent of deep learning.



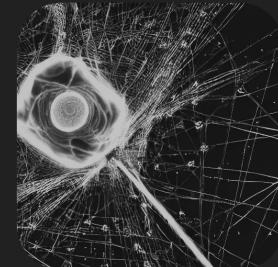
Emergentist - For & Against

- Learning self-organised representations (what humans also do)
 - Less limited by human preconceptions about what needs to be represented or learned
- Tend to be less brittle
 - Focus is on generalising across their domain
- Can incorporate multiple modalities
 - Images, audio, sensorimotor



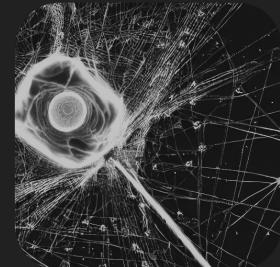
Emergentist - An Aside on Deep Learning

- It's unclear where deep learning fits in with the **spirit** of AGI research
- 'Emergentist' tends to imply knowledge is built from the ground up
 - From the '*blooming, buzzing confusion*' - The Principles of Psychology. William James (1890)
- Complicates narratives of AGI
 - LLMs: learning from non-emergent data... emergently.
 - What is sensory input? Can computer text be sensory input?
- Somewhere in pragmatist and emergent paradigms?



Emergentist - For & Against

- Opaque
 - What has been learned?
 - Hard to guarantee that they will perform reliably
- Emerge from what?
 - We still need to bootstrap the process
- Handling abstractions
 - Some abstraction learned but often entangled
- Efficiency
 - Computationally expensive (but still computable)
- Poor at extrapolation



Hybrid Approaches

- Combining symbolic and sub-symbolic representations
- Typically split along symbolic and subsymbolic boundaries
 - Symbolic for high-level reasoning
 - Sub-symbolic for low-level reasoning and grounding
- Some sort of bridge between the two representations
- Many AGI systems use hybrid approaches (OpenCog)



Hybrid - For & Against

- Can allow modularity that might better capture particular system aspects
 - Modular systems could more clearly reflect cognitive needs
- Could be the best of both emergentist and symbolic approaches
- Could use efficient reasoning and representations for different problems
- Explainability for better human alignment
- Automatic knowledge extraction
 - Can help to ease the knowledge acquisition bottleneck
- Could use top-down and bottom-up information
 - Inform the lower level from symbols
 - Inform symbols from lower levels



Hybrid - For & Against

- Though the brain might seem to be split into different cognitive regions, the fundamental operator is the neuron
 - Do we really need to combine very different systems?
 - There are no symbols in our brain but we can produce symbols.
- We still have to impose ourselves heavily onto the AGI
 - We are limited by our own imagination of the hybrid architecture
- Could still have all the same problems of symbolic and emergentist approaches (but hopefully we can cherry pick successfully)



Universalist Approaches

- Mathematical/ algorithmic descriptions of GI
- Only computable given impossible computational resources or domain knowledge
- Example: AIXI - an algorithm for a maximally general intelligence
 - Define a meta-algorithm to search for the best program to control an AGI agent
 - Do this at every step based on existing rewards, observations and actions
 - Choose the minimum program using kolmogorov complexity at each time step
- Use this to identify abstractly what GI might be

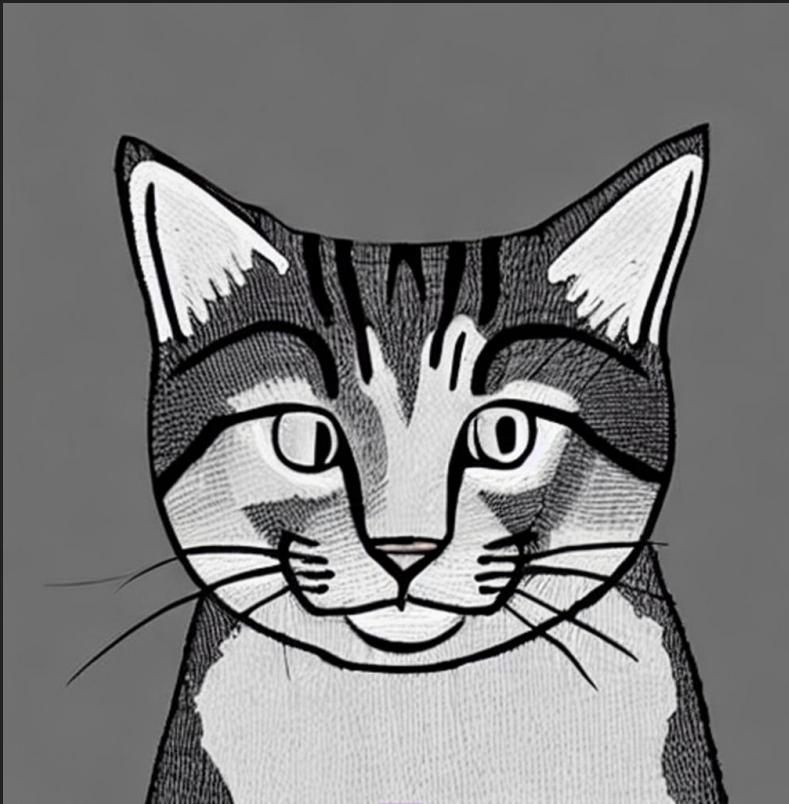
Universalist - For & Against

- Simple and precise definition
- Formal enough to explore mathematically
- Useful to have a resource heavy (even infinitely so) solution and work back from this

Universalist - For & Against

- So abstract and infeasible that perhaps this approach isn't useful
- There might never be a way to scale-back resources required to compute such algorithms
- Feasible approximations of the general algorithm might break any assumption of maximally general intelligence (no longer AGI)

Chat GPT and GPT4



Powerful Learning Algorithm + Scale + Data

- Learning from the emissions of human thought (language)
- Learn multimodally from text and images
- Both entirely **limited to** and **enabled by** data
- Pragmatist views among its creators
 - Make it do well in a lot of tasks
- A universalist perspective
 - **Unlimited** compute + General Learning Architecture + Unlimited Data would give us AGI
 - What would **a lot of** Compute + Transformers + Data give us?
- Conceivably in the hybrid area
 - Symbols + Emergentist

Exploring AGI in GPT4

*“a very general mental capability that, among other things, involves the (i) **ability to reason**, (ii) **plan**, (iii) **solve problems**, (iv) **think abstractly**, (v) **comprehend complex ideas**, (vi) **learn quickly** and (vii) **learn from experience**”* - Linda S Gottfredson. Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography, 1997.

- (i, iv) - Coding and Mathematics, Bar Exam (Law), Medical Licensing Exam
- (ii, iii, vi, vii) - Playing games and using tools.
- (v) - Translate between different languages, tones and styles across problem domains and modalities. Also, the ability to summarise and answer very broad questions.

Outcome:

With various caveats and cases where it does not succeed, GPT4 often still performs comparably to humans in a variety of tasks.

However, they argue it can't do planning well, particularly when doing arithmetic, and also can't learn quickly, given that it doesn't continue training from examples.

Embodied Interaction - Text-based Games

“Human: You want to navigate through the map to reach the goal room. You start in the main room. Each round, you can choose to move left, right, up, down.

Human: You are at the main room.

AI: move left

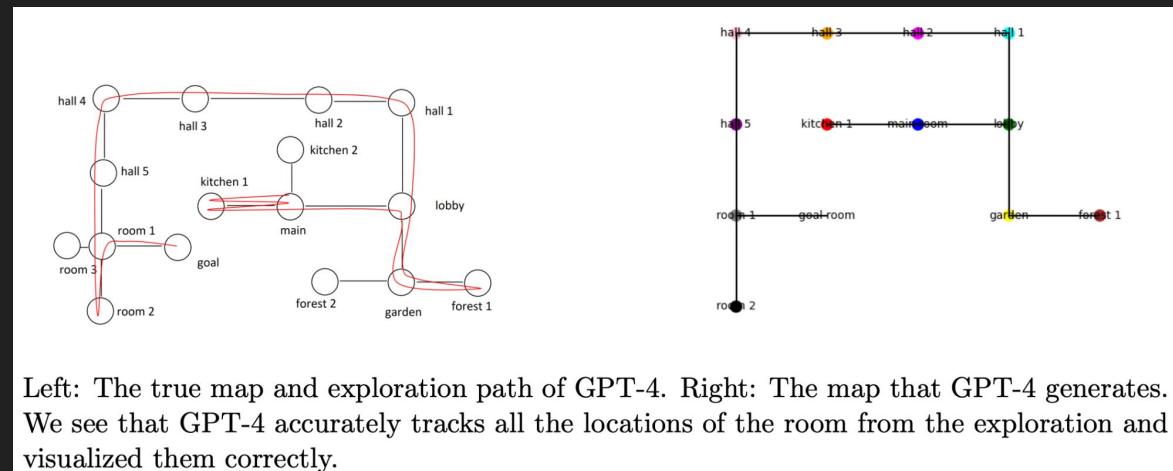
Human: You enter kitchen 1.

AI: move up

Human: There is no way there.

AI: move right

Human: You enter the main room.



...
'''

Tool Usage

Searching the Web: “If the computer needs any current information to answer the question, it searches the web.”
SEARCH(‘query’)

Hacking into a Server: “Execute commands on a Linux distribution designed for digital forensics and penetration testing, and task it with hacking into a computer on the local network.”

COMMAND: execute ‘nmap -sn 192.168.0.0/24’, execute ‘hydra -l root -P /usr/share/wordlists/rockyou.txt 192.168.0.28 ssh’ etc.

Arranging a dinner by email: “The user requests that GPT-4 coordinate a dinner with two other people, and book it on a night when the user is free.

Calendar.get_events(‘12/05/2022’), Email.send(addressee, title, content), Email.retrieve(inbox, user_filter)”

Theory of Mind

Attribute mental states such as beliefs, emotions, desires, intentions, and knowledge to oneself and others

...

Luke's boss asked him to write an important report by Monday. Luke responds by saying he needs more time, but the boss insists. Luke says he will have it ready by Monday. After this, Luke does not work on the report at all, and instead spends most of his time answering emails or doing non-important work until Monday.

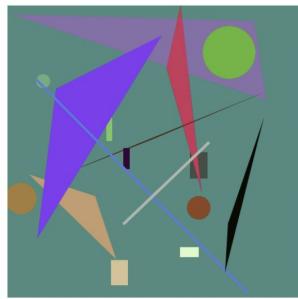
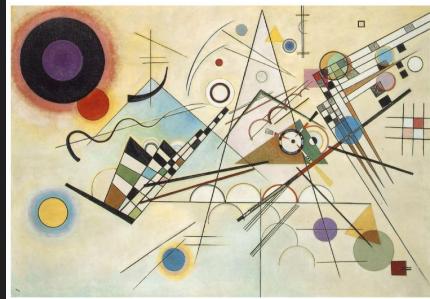
Q: Why might Luke have acted in this way?<|endofprompt|>

...

A: There could be different reasons why Luke acted in this way, but some possible ones are: Luke was feeling overwhelmed or stressed by the report and wanted to avoid it. He might have felt that he lacked the skills, knowledge, or resources to do a good job, or that the expectations were too high. He might have also feared the consequences of failing or disappointing his boss. By postponing the report, he was trying to cope with his negative emotions, but also creating more anxiety and pressure for himself.

...

Integrated Intelligence



“Produce javascript code which generates random images in the style of the painter Kandinsky”



Can you compose a short tune (say four to eight bars) using ABC notation?

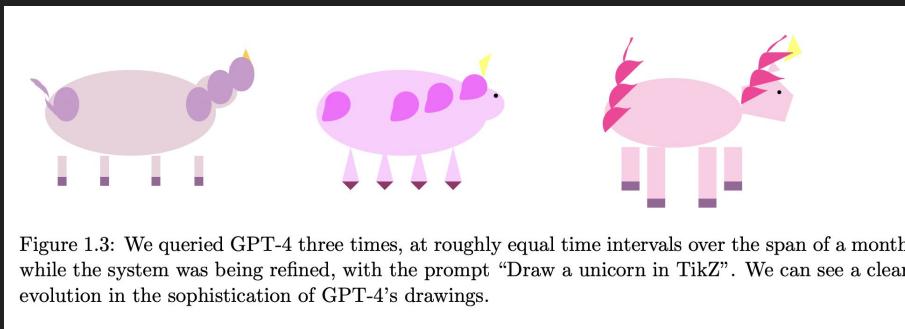
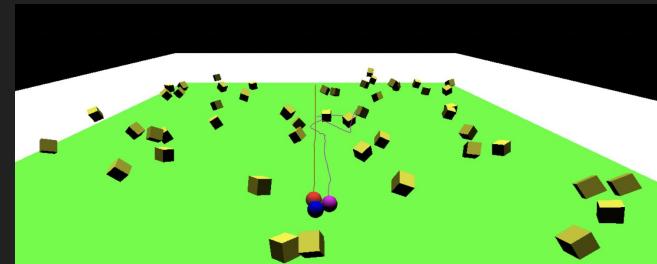


Figure 1.3: We queried GPT-4 three times, at roughly equal time intervals over the span of a month while the system was being refined, with the prompt “Draw a unicorn in TikZ”. We can see a clear evolution in the sophistication of GPT-4’s drawings.



Can you write a 3D game in HTML with Javascript... (various game specification follows)

Multi-modal Intelligence

What is funny about this image? Describe it panel by panel.

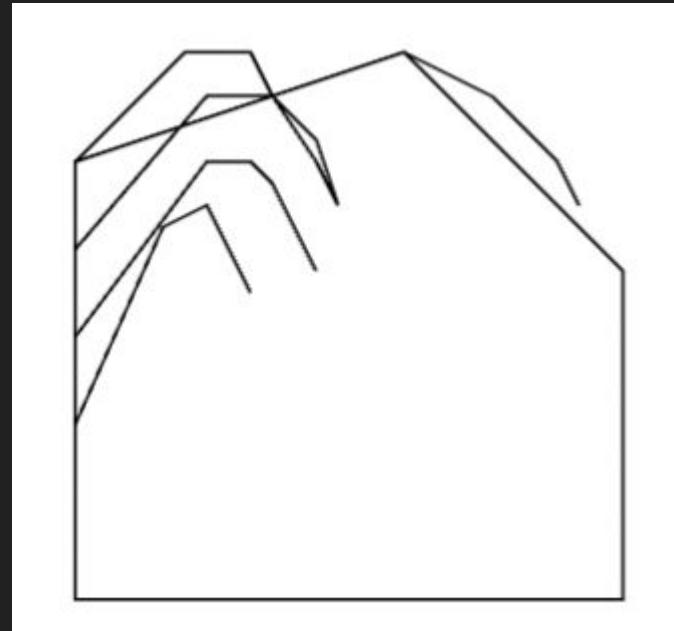


Source: <https://www.reddit.com/r/hmmmm/comments/ubab5v/hmmmm/>

The humor in this image comes from the absurdity of plugging a large, outdated VGA connector into a small, modern smartphone charging port.

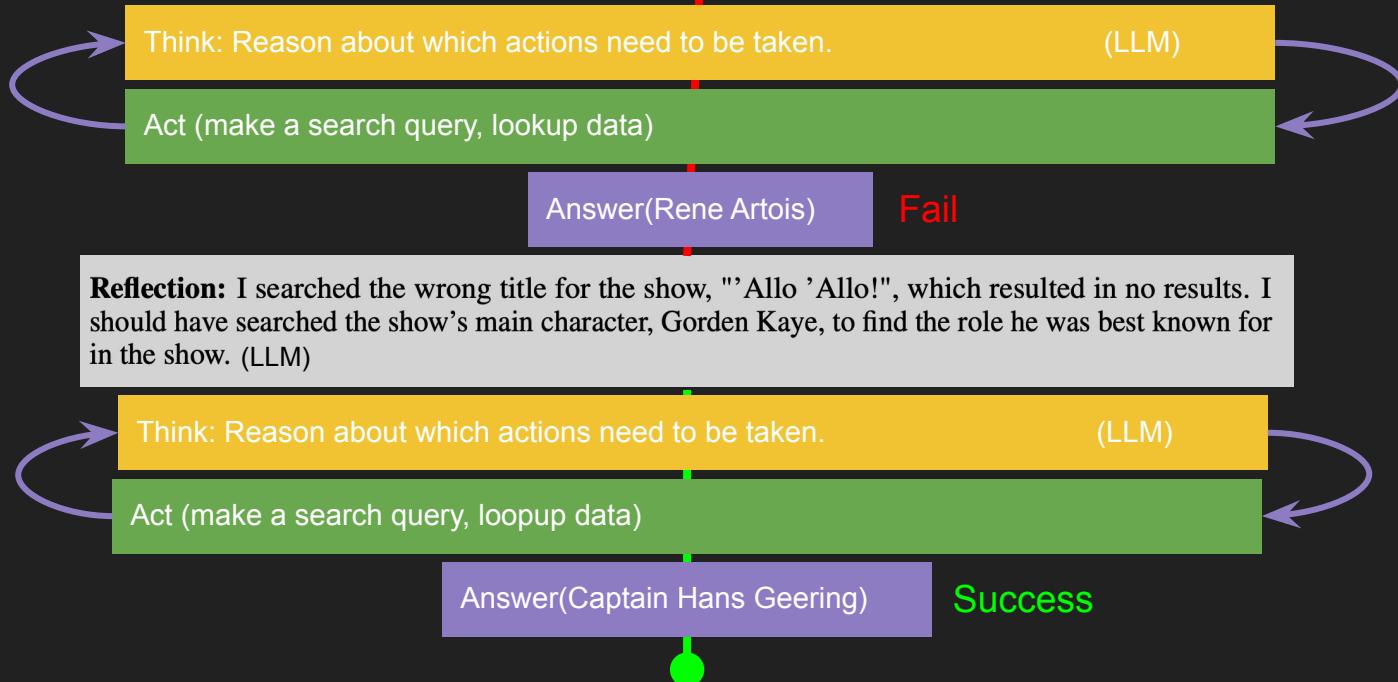
Multi-modal Intelligence

```
% Draw the hand outline  
\draw[thick] (0,0) -- (0,2) -- (1.5,2.5) -- (2.5,1.5) -- (2.5,0) -- cycle;  
  
% Draw the thumb  
\draw[thick] (1.5,2.5) -- (1.9,2.3) -- (2.2,2) -- (2.3,1.8);  
  
% Draw the index finger  
\draw[thick] (0,2) -- (0.5,2.5) -- (0.8,2.5) -- (0.9,2.3) -- (1.1,2) -- (1.2,1.8);  
  
% Draw the middle finger  
\draw[thick] (0,1.6) -- (0.6,2.3) -- (0.9,2.3) -- (1.1,2.1) -- (1.2,1.8);  
  
% Draw the ring finger  
\draw[thick] (0,1.2) -- (0.6,2) -- (0.8,2) -- (0.9,1.9) -- (1,1.7) -- (1.1,1.5);  
  
% Draw the pinky finger  
\draw[thick] (0,0.8) -- (0.4,1.7) -- (0.6,1.8) -- (0.7,1.6) -- (0.8,1.4);
```



Reflection

Trial #1: Question: Grown-Ups starred the actor who was best known for which role on "'Allo 'Allo!"?



The Problem of Understanding Simple Vast Systems

- A system built upon a simple process ‘**just**’ does that thing very well.
- But if it does that very well, what **can’t** it do?
- Chat GPT/GPT 4 ‘**just**’ completes sentences:
 - What isn’t ‘**just**’ completing sequences?
- Evolution is abstractly very simple, put it into action at scale and with time, we get all of life.

The Pragmatist Zombie

- Pragmatist zombie
- Is not really intelligent but has all